STIC-ILL QL366.J6

From: Huynh, Phuong N.
Sent: Tuesday, June 18, 2002 1:15 PM
To: STIC-ILL
Subject: RE: 09/761,636

Please deliver the following:

Genomics 42: 483-488; 1997

J Biochem 274: 32127; 1999

3

J Eurkaryot Microbiol 43: 303-314; 1996

Thanks,
Neon
Art unit 1644
Mail 9E12
Tel 308-4844

21. Huyunh, T. U., Young, R. A. & Davis, R. W. 1985. Screening of cDNA libraries. *In:* Glover, D. M. (ed.), DNA Cloning: A Practical Approach. IRL Press, Oxford, UK. 1:49–78.

22. Hunziker, W., Spiess, M., Semenza, G. & Lodish, H. 1986. The sucrase–isomaltase complex: primary structure, membrane-orientation, and evolution of a stalked, intrinsic brush border protein. *Cell*, **46**:227–234.

23. Jeffrey, P. L., Brown, D. H. & Brown, B. I. 1970. Studies of lysosomal α-glucosidase. I. Purification and properties of the rat liver enzyme. *Biochemistry*, **9**:1403–1415.

24. Jeffrey, P. L., Brown, D. H. & Brown, B. I. 1970. Studies of lysosomal α-glucosidase. II. Kinetics of action of the rat liver enzyme. *Biochemistry*, **9**:1416–1423.

25. Kaplan, A., Achord, D. T., & Sly, W. S. 1977. Phosphohexosyl components of a lyososomal enzyme are recognized by pinocytosis receptors on human fibroblasts. *Proc. Natl. Acad. Sci.* USA, **74**:2026–2030.

26. Katoh, M., Hirono, M., Takemasa, T., Kimura, M. and Watanabe, Y. 1993. A micronucleus-specific sequence exists in the 5'-upstream region of calmodulin gene in *Tetrahymena thermophila*. *Nucl. Acids Res.*, **21**:2409–2414.

27. Koster, J. F. & Slee, R. G. 1977. Some properties of human liver acid α-glucosidase. *Biochim. Biophys. Acta*, **482**:89–97.

28. Lutcke, H. A., Chow, K. C., Mickel, F. S., Moss, K. A., Kern, H. F. & Scheele, G. A. 1987. Selection of AUG initiation codon differs in plants and animals. *EMBO J.*, **6**:43–48.

29. McKnight, S. L. and Kingsbury, R. 1982. Transcriptional control signals of a eukaryotic protein-coding gene. *Science*, **217**:316–324.

30. Müller, M. 1972. Secretion of acid hydrolases and its intracellular source in *Tetrahymena pyriformis*. *J. Cell Biol.*, **52**:478–487.

31. Palmer, T. N. 1971. The substrate specificity of acid α-glucosidase from rabbit muscle. *Biochem. J.*, **124**:701–711.

32. Palmer, T. N. 1971. The maltase, glucoamylases and transglucosylase activities of acid α-glucosidase from rabbit muscle. *Biochem. J.*, **124**:713–724.

33. Rosenfeld, E. L. 1975. α-Glucosidase (γ-amylases) in human and animal organisms. *Pathol. Biol.*, **23**:71–84.

34. Rosenfeld, M. G., Kreibich, G., Popov, D., Kato, K. & Sabatini, D. D. 1982. Biosynthesis of lysosomal hydrolases: their synthesis in bound polysomes and the role of co- and post-translational processing in determining their subcellular distribution. *J. Cell. Biol.*, **93**:135–143.

35. Sambrook, J., Fritsch, E. F. & Maniatis, T. 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

36. Sato, K., Inazu, A., Yamaguchi, S., Nakayama, T., Deyashiki, Y., Swada, H. & Hara, A. 1993. Monkey 3-deoxy glucosone reductase: Tissue distribution and purification of three multiple forms of the kidney enzyme that are identical with dihydrodiol dehydrogenase, aldehyde reductase and aldolase reductase. *Arch. Biochem. Biophys.*, **307**:286–294.

37. Sly, W. S. & Fisher, H. D. 1982. The phosphomannosyl recognition system for intracellular and intercellular transport of lysosomal enzymes. *J. Cell. Biochem.*, **18**:67–85.

38. Tabas, I. & Kornfeld, S. 1980. Biosynthetic intermediates of β-glucuronidase contain high mannose oligosaccharides with blocked phosphate residues. *J. Biol. Chem.*, **255**:6633–6639.

39. Taniguchi, T., Mizuochi, T., Banno, Y., Nozawa, Y. & Kobata, A. 1985. Carbohydrates of lysosomal enzymes secreted by *Tetrahymena pyriformis*. *J. Biol. Chem.*, **260**:13941–13946.

40. von Figura, K. & Hasilik, A. 1980. Lysosomal enzymes and their receptors. *Ann. Rev. Biochem.*, **55**:167–193.

41. von Figura, K. & Klein, U. 1979. Isolation and characterization of phosphorylated oligosaccharides from α-N-acetylglucoaminidase that are recognized by cell-surface receptors. *Eur. J. Biochem.*, **94**:347–354.

42. von Heijne, G. 1986. A new method for predicting signal sequence cleavage sites. *Nucl. Acids Res.* **14**:4683–4690.

J. Euk. Microbiol., 43(4), 1996, pp. 303–313

Molecular Characterization of the D Surface Protein Gene Subfamily in *Paramecium primaurelia*

FLORENCE M. BOURGAIN-GUGLIELMETTI[1] and FRANCOIS M. CARON*[2]

**Laboratoire de Génétique Moléculaire, Ecole Normale Supérieure, 46 rue d'Ulm, 75230 Paris, Cedex 05, France*

ABSTRACT. When *Paramecium primaurelia* expresses the D serotype, a major high molecular weight mRNA species is detected in the cytoplasm. Using the cDNA derived from this mRNA as a probe, three very similar genes, Dα, Dβ and Dγ, were cloned. Of these three genes, we show that only the Dα mRNA is present in the cytoplasm of cells expressing the D serotype and corresponds to the major mRNA species. The nucleotide sequence of the entire coding region of the Dα gene, as well as the upstream and downstream sequences, has been determined. The 7632-nucleotide open reading frame encodes a putative protein that displays the characteristic cysteine residue periodicity of *Paramecium* surface antigens but does not contain central tandemly repeated sequences. Partial sequences of the two nonexpressed genes Dβ and Dγ indicate a high percentage of identity (90%–95%) with the Dα gene, suggesting that Dβ and Dγ genes are either very similar surface protein genes whose transcription is repressed trough mutual exclusion, or perhaps are pseudogenes. A region of variable DNA rearrangement was identified 1 kb upstream of the Dγ gene. This macronuclear region arises from the same micronuclear locus by alternative excision of internal eliminated sequences during macronuclear development.

Supplementary key words. Alternative DNA rearrangements.

PARAMECIUM primaurelia possesses a family of surface antigen genes. In most cases, only one of these genes is expressed (exclusion rule) and the corresponding protein covers the external surface of the cell constituting the cell coat. Two surface antigens called G and D have been extensively studied by biochemical, genetical and immunological means (for a review see [29]). The G protein is stably expressed in the 15° C–28° C temperature range whereas the D protein is expressed above 30° C. At the molecular level, the G protein gene has been cloned and entirely sequenced from two geographically distinct isolates: strains 156 and 168 [26, 27]. They are huge allelic proteins with molecular masses in the 275 kDa range and display along their whole sequence a similar pseudoperiodic structure. This pseudoperiodic structure is shared by all se-

[1] Present address: *Institut de Génétique Moléculaire, Unité 301 INSERM, 27 rue Juliette Dodu, 75010 Paris, France.*

[2] To whom correspondence should be addressed. Telephone: 33-144-3239-48; Fax: 33-144-3239-41; Email: caron@wotan.ens.fr

quenced *Paramecium* surface antigen genes. At the center of the aminoacid sequences of these two surface antigens, 156G and 168G, four almost perfect repeats of about 70 residues are present and appear as a distinctive feature of these proteins: indeed, the similarity between the two allelic sequences is extremely high (98%) except for these central repeats where the similarity percentage drops to 60% [8]. Different immunological arguments indicate that these four central repeats form four identical domains, which are the only parts of the molecule accessible from the external medium [3, 5, 26].

Paramecium primaurelia shares the common characteristic of ciliates: nuclear dimorphism. In each cell, two types of nuclei coexist: a micronucleus, which is essentially transcriptionally inactive and acts as a germinal nucleus; and a macronucleus, which is transcriptionally active and acts as a somatic nucleus. After each sexual process, a new macronucleus is made from a copy of the zygotic nucleus and the old macronucleus is destroyed. A complex DNA rearrangement process takes place during the biogenesis of this macronucleus consisting in three types of modifications: internal sequence elimination, chromosome fragmentation and DNA amplification. Different types of eliminated micronuclear sequences have been reported in the literature. Among them, one type called IES (internal eliminated sequence) are unique DNA sequences rich in A and T bases and bordered by two direct repeats of the dinucleotide 5'-TA-3'. They are eliminated by an excision-religation mechanism that maintains only one of the two direct repeats. Although recent efforts have been made to determine the nature of this process (cis specific sequences, trans effecting protein factors), it remains essentially unknown (for a review, see [33]).

In this paper, we have tried to extend the type of structural research already carried out on the G surface protein of strain 156 to the D surface protein of the same strain which, as mentioned above, is expressed at higher temperature. Using the mRNA sequence of the D protein as a probe, we have screened a genomic macronuclear library and cloned three genes whose sequences cross-hybridize strongly with the mRNA sequence. The three genes have been mapped and found to have extremely similar sequences. Only one of the genes is expressed in the D serotype. The complete nucleotide sequence of the expressed gene has been determined and from the deduced amino acid sequence we show that the structure of the D protein resembles that of the G protein except for the absence of the central repeats; a surface protein of *Paramecium tetraurelia,* 51C, has also been reported to lack the central repeats [25]. During the course of this work, an unusual variable region was found close to the 5' end of the coding sequence of one of the two nonexpressed genes. Close inspection of this upstream region shows that it varies from one macronuclear copy to another. The sequences of some of these various macronuclear versions have been determined and we suggest that they could be generated by alternative elimination of IES.

MATERIALS AND METHODS

Cell line and cultivation. Cells from *Paramecium primaurelia* wild-type strain 156 were grown in "scotch grass" infusion as described by Sonneborn [38], bacterized the day before use with Klebsiella pneumoniae and supplemented with 0.8 μg/ml of β-sitosterol (Merck). Cultivation were carried out at 24° C or 33° C for expression of the G or D serotypes respectively. Surface antigen expression was determined and routinely checked during culture expansion by the immobilization test [2] using antisera raised against either 156G or 156D surface antigens. Cultures were used when 100% of the cells expressed a given surface antigen. Cells were collected by centrifugation and washed in Dryl's solution [12]. The compact pellet of cells was either used immediately for DNA extraction or frozen for RNA extraction. In the latter case, pelleted cells were poured dropwise into liquid nitrogen and kept at −80° C.

DNA analysis. Genomic *Paramecium* DNA was isolated as described previously in Prat [26]. Electrophoresis and Southern blot hybridization were performed according to usual methods [35].

RNA extraction and analysis. Two methods were used to prepare total RNA from frozen pelleted *Paramecium* cells: the first method was adapted from Chirgwin et al. [9] and described in Meyer et al. [24]. RNA used in the experiment of Fig. 1 was prepared by this method. The second method was method 2 described in Meyer et al. [24] with minor modifications. This method was used for the preparation of total RNA in the experiments of Fig. 5.

RNA were run on methylmercury gels and blotted on Hybond N^+ membranes according to Sambrook [35]. Northern blots were hybridized with oligonucleotides Osr1, Osr2 and Osr3 in the buffer described by Emilsson and Kurland [13].

Stringency determination for oligonucleotide hybridization. To find conditions suitable for the specific hybridization of the three oligonucleotides Osr1, Osr2, Osr3 (see Fig. 4 for their sequences) with the corresponding identical sequences, we immobilized on Nylon$^+$ membranes using a dot blot apparatus, 2 ng to 2 μg of recombinant plasmid DNA containing as insert either S1 or S2 or S3 (Fig. 2 for the localization of these fragments). These sequences contained respectively the Osr1, Osr2 and Osr3 sequences. DNA were denatured in 0.4 M NaOH for 30 min. at 65° C prior to filter binding. The filters were successively probed with each oligonucleotide and washed in 0.2× SSC (150 mM NaCl, 15 mM Na Citrate pH 7), 0.5% SDS at increasing temperatures for 30 min. At low stringency (45° C) the three oligonucleotides hybridized to each of the plasmids. Above 60° C, the 23-bp oligonucleotide Osr1 hybridized with S1 but not with S2 or S3. Above 55° C, the two 23-bp oligonucleotides Osr2 and Osr3 hybridized with the fragment containing the exact sequence and not with the two others. Thus, by using these stringency conditions, oligonucleotides Osr1, Osr2 and Osr3 constitute specific probes for genes Dα, Dβ and Dγ, respectively.

Cloning of probe pEDx. The method used was described in Meyer et al. [24]. A short account is given here: polyA$^+$ mRNA were purified from G or D expressing cells on an oligo-dT cellulose column (Fig. 1A), denatured with methyl mercury and sized-fractionated on a sucrose gradient. cDNA was made from selectively enriched high molecular weight mRNA and used as probe in a differential screening of an EcoRI library. Plaques positive with the cDNA D probe and negative with the cDNA G probe were selected, subcloned and their phage DNA purified and analyzed by restriction mapping. An EcoRI fragment of 1.6 kb was subcloned in pUC18. This recombinant plasmid (pEDx) was previously referred to as pED1 when subcloned in pBR328 [24]. Used as a probe on a Northern blot of polyA$^+$ mRNA of G or D expressing paramecia (Fig. 1) pEDx specifically hybridizes with the intense band and not with the weak band in polyA$^+$ mRNA from D expressing cells (see text and Fig. 1B, lane 2).

Genomic library construction. Total *Paramecium* DNA was partially digested with EcoRI and DNA fragments in the 15- to 25-kb range were size-fractionated on a low melting agarose gel and purified by the agarase method prior to ligation with λ EMBL3 vector arms. Thirty thousand plaques were screened with pEDX and 82 positive plaques were selected at random without any size discrimination and purified by another round of hybridization.

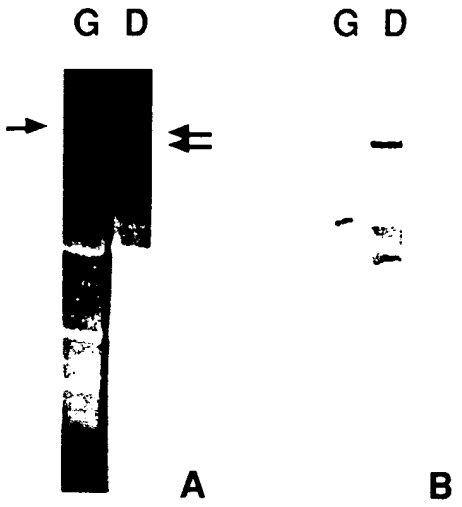

Fig. 1. mRNA characterization of G and D surface antigens. A. Methyl mercury RNA gel analysis of polyadenylated RNA extracted from G or D expressing cells. The abundant high molecular weight surface antigen mRNAs are visible by ethidium bromide staining and are indicated by arrows. Based on the size of the two ribosomal RNAs, the size of the G mRNA is estimated to be 8000 bases, those of the two mRNAs specifically expressed in serotype D 8000 bases and 7500 bases. B. The same gel blotted and probed with pEDx showing the specificity of this probe for the D mRNA.

Isolation of recombinant clones and DNA sequencing. DNA was purified from recombinant phages according to Sambrook [35]. The restriction maps of recombinant phages was determined by the cos mapping technique [34].

Restriction fragments to be sequenced were subcloned into pUC19. The Promega Erase a Base kit was used to create a nested set of deletions with exonuclease III (Promega, Madison, WI, USA). The resulting plasmids were transformed into *E. coli* strain MR32 to produce DNA for standard double stranded sequencing. Sequencing reactions were performed using the sequenase DNA sequencing kit version 2.0 (USB, Cleveland, OH, USA). In general, only one strand was sequenced; but, each time there was an ambiguity in the sequence determination, both strands were sequenced. For the contiguous sequences produced after exonuclease III deletions, both strands of the overlapping segments were sequenced.

In region A, the EcoRI restriction fragments of 1.2, 1.6 (EDx), 1.6, 4.3, 2.1 and 3.1 kb from phage λD2 bearing the 156Dα gene (Fig. 2) were entirely sequenced. For the 156Dβ gene, the HindIII restriction fragments of 3.8, 1.3, 0.8, 0.2 and 5.6 kb (Fig. 2) and the 1.4-kb EcoRI restriction fragment of λD24 (Fig. 2) were partially sequenced from their extremities. In the B region, the extremities of λD81 EcoRI restriction fragments of 5.7, 0.9, 1.6 and 8.9 kb were also sequenced. Each of the HincII-EcoRI fragments of phages λD81, λD19, λD57, λD22 and λD55 were subcloned and entirely sequenced, but only the extremities of the corresponding fragment from λD15 was sequenced.

DNA and protein sequences were analyzed using the University of Wisconsin GCG sequence analysis software package [11] and DNA Strider [19].

DNA sequence accession numbers. The DNA sequences obtained were submitted to both the EMBL Nucleotide Sequence Database and GenBank. Unless otherwise noted, database accession numbers are listed with the EMBL accession number first, followed by the GenBank Accession number in parenthesis.

PCR amplification. The 25 μl reaction mixtures containing 10–100 ng of genomic DNA or 0.02–20 ng of λ recombinant DNA for control were amplified in a Perkin Elmer Cetus apparatus for 32 cycles (92° C, 1 min.; 65° C, 1 min. 15 s; 72° C, 1 min. 30 s) and with a final extension time of 10 min. at 70° C. The amplified products were directly used for electrophoresis or after purification with Qiaquick Spin kit (250) (Qiagen, Chatsworth, CA, USA) to eliminate oligonucleotides of less than 30 bases.

RESULTS

Cloning of a probe of the D antigen subfamily. Figure 1A represents an ethidium bromide stained RNA gel of G or D expressing cells. In both cases, a major high molecular weight band corresponding to the abundant mRNA of the expressed surface antigen is detectable [24]. The mRNA from G expressing paramecia consists of a single band, which has been shown to contain only one mRNA species [23, 24], whereas the mRNA from D expressing cells consists of two bands, one intense band, which migrates slightly faster than the corresponding G band; and a weaker band, which has the same mobility as the G band. To obtain probes of the G and D surface antigens, we took advantage of the fact that the mRNA are polyadenylated and of high molecular weight [24, 31]: mRNA were purified on an oligo-dT cellulose column and size-fractionated on a sucrose gradient [24]. Radioactive cDNA were prepared from the fractions enriched in surface antigen mRNA with an oligo-dT primer and used as probes in a differential screening of an EcoRI library. Plaques positive with the D probe and negative with the G probe were selected, subcloned, and their phage DNA purified and analyzed by restriction mapping. An EcoRI fragment of 1.6 kb, pEDx (Fig. 2), common to all these recombinants, was subcloned in pUC18 (the same fragment cloned in pBR328 was already mentioned in previous publications as pED1: [23, 24]) and used to probe a Northern blot of polyA$^+$ RNA of G or D expressing paramecia (Fig. 1B). It hybridizes specifically to the intense band of D expressing cells and not to the weak band. A G probe cloned in the same way hybridizes only to the unique G specific band [24]. This indicates that the weak band from D expressing cells that migrates with the same electrophoretic mobility as the G band differs in sequence from the 3′ region of mRNA of the intense band and from the G mRNA. No attempt has been made to characterize the molecular species contained in this weak band since, if it is the mRNA of a coexpressed surface antigen, it is likely that it does not belong to the D subfamily.

Cloning and characterization of the D antigen subfamily. Since surface antigen mRNA molecules are long molecules, we needed recombinants with large inserts to cover the whole gene(s). An EcoRI library was constructed from a partial digestion of *Paramecium* DNA after size selection of fragments in the 15- to 25-kb range and insertion in the λ EMBL3 vector. Using pEDX as probe, 82 positive plaques were selected and subcloned, and 30 randomly chosen were analyzed by restriction mapping. All these phages can be associated to two genomic regions A and B (Fig. 2): 24 phages originate from region A and six from region B. Region A (top of Fig. 2) is more than 30-kb long and is entirely covered by four phages (out of the 24): λD1, λD2, λD3 and λD24. The first three contain an EcoRI fragment S1 (indicated by a thick line in Fig. 2) which, by restriction mapping and DNA sequencing, is identical to pEDX. Phage λD24 contains a HindIII–EcoRI restriction fragment called S2, which is also represented by a thick band in Fig. 2. S2 is the smallest fragment we could find that strongly hybridizes with pEDX (Fig. 2). The absence of EcoRI sites in this S2 fragment indicates that S2 is similar but not identical to pEDX. S1 and

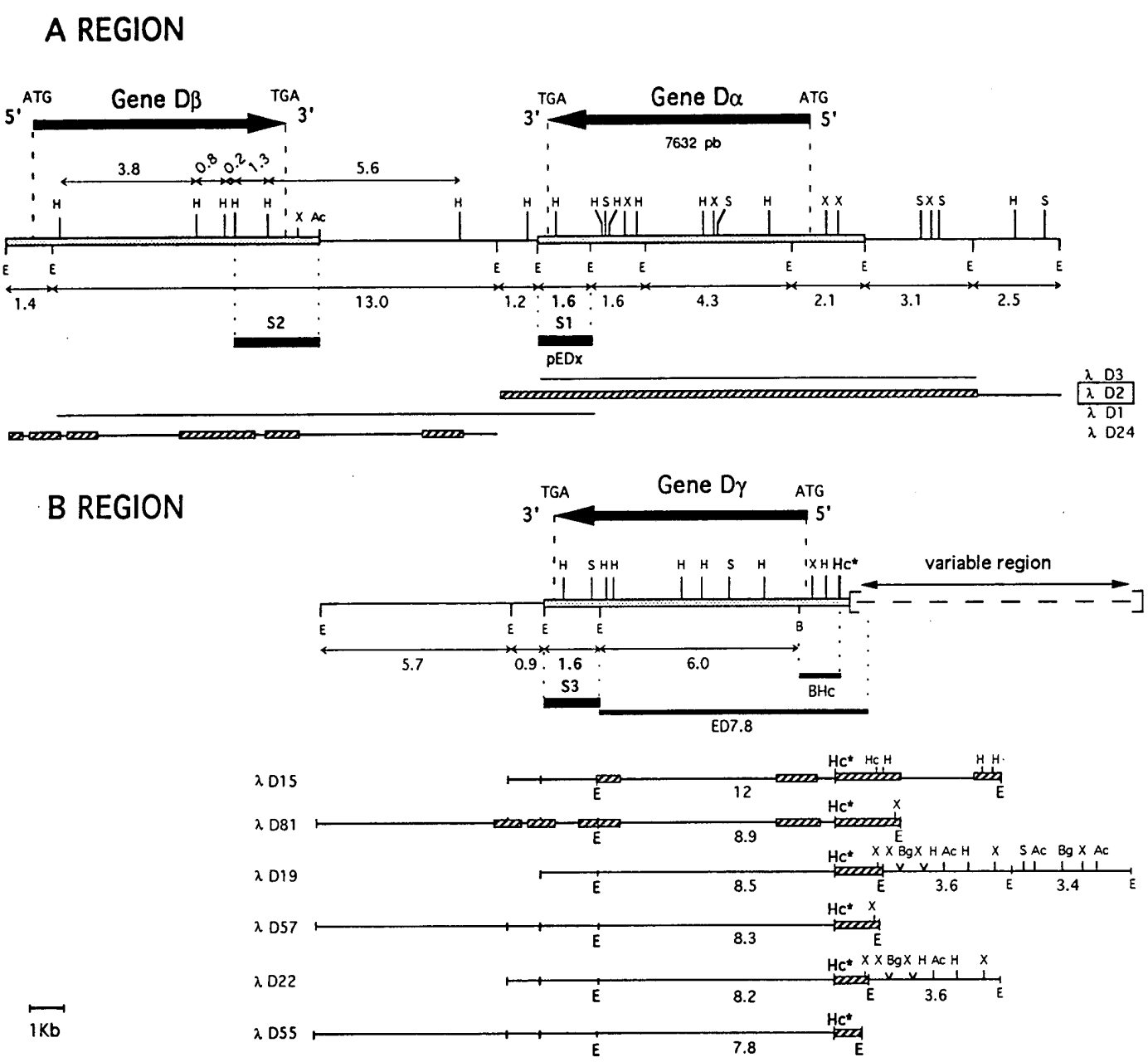

Fig. 2. Restriction mapping of the regions containing the D genes. The two maps of genomic regions A (top) and B (bottom) are shown. Restriction sites are: Ac, AccI; Bg, BglII; E, EcoRI; Hc, HincII; H, HindIII; S, ScaI; X, XbaI. The inserts of phages are shown below the maps. λD1, λD2, λD3 and λD24 cover the A region; λD15, λD81, λD19, λD57, λD22 and λD15, the B region. Three fragments, S1, S2 and S3, shown by thick lines at the 3' end of genes Dα, Dβ and Dγ, respectively, hybridized with probe pEDx. A rough estimate of the extent of similarity between the three genes is represented by the dotted thick lines. The black thick arrows indicate the sequenced Dα gene and the two partially sequenced Dβ and Dγ genes. The two ends of each black arrow are located at the initiator codon ATG and at the stop codon of the coding regions. All sequenced regions are indicated by thick hatched lines on the maps of the phages from which the corresponding restriction fragments used for sequencing have been subcloned. The sizes of EcoRI or HindIII restriction fragments are given under each fragment. Probes used in this work: pEDX, BHc, ED7.8 are indicated by thick lines under the corresponding fragment except for probe ED7.8, which corresponds to the 7.8-kb EcoRI fragment of phage λD55. The right part of the B region map is variable in size due to alternative DNA rearrangements.

S2 are contained in λD1 showing their proximity on the genome. The large size of region A (> 30 kb) and the presence of two similar sequences strongly suggest the presence of two genes in region A and, indeed, we shall show (see further in the text) by DNA sequencing that two similar D genes, Dα gene on the right and Dβ gene on the left, are present in inverted orientations, as shown in Fig. 2 by the two large arrows.

The restriction maps of the six phages related to region B are displayed at the bottom of Fig. 2. An EcoRI fragment of 1.6 kb, called S3, is present in all phages and has a sequence similar but not identical to pEDX since, for instance, pEDX does not contain a ScaI site (S site in Fig. 2). All these phages contain an HincII site designated by Hc*. At the left of this site, the restriction maps of these phages correspond to a unique map. On

the right of the Hc* site, the maps differ (see for instance in Fig. 2 the distances between the Hc* site and the next EcoRI site on the right). The identity of the maps at the left of site Hc* suggests they originate from the same genomic region; an unlikely alternative explanation would be the existence of a region duplicated several times and maintained identical. The common genomic origin of the six phages is further supported by EcoRI restriction mapping of the region around the Hc* site by Southern blot (Fig. 3, 11), which reveals the presence of a 7.5- to 9-kb smeared band and of a 12-kb band corresponding to region B. Each of the five phages λD81, λD19, λD57, λD22 and λD55 has an EcoRI fragment containing the Hc* site whose size is in the 7.8- to 8.9-kb range (8.9, 8.5, 8.3, 8.2 and 7.8 kb, respectively) (Fig. 2), whereas λD15 has a 12-kb EcoRI fragment containing the Hc* site. This shows that the B region is heterogeneous and complex and that the six phages provide a good albeit incomplete representation of it. For two phages, λD19 and λD22, the extreme right part of their maps is again identical (a common 3.6-kb EcoRI fragment). Therefore, the right part of the maps might reflect variable DNA rearrangements during macronuclear DNA differentiation.

The next step was to determine whether or not the regions of similarity represented by S1, S2 and S3 were limited to these sequences. For this purpose, we used the series of EcoRI fragments of phage λD2 to probe different DNA digests of the other phages. The results (not shown) can be interpreted in the following way: from each of the S1, S2 and S3 fragments, an 8- to 9-kb region of similarity can be determined which, as shown later, covers three genes and contains S1, S2 and S3 at one end (dotted thick lines in Fig. 2). Outside of these regions, the similarity drops. For instance, the two EcoRI fragments (1.2- and 3.1-kb long) (Fig. 2) that frame gene Dα do not hybridize to the corresponding parts of genes Dβ and Dγ. They are unique sequences in the genome. Since pEDX (which is identical to S1) contains the 3′ end of the gene because of the cloning process, the distal positions of S1, S2 and S3 with respect to the three regions of similarity must represent the 3′ ends of the genes.

How many genes in the D subfamily? Three putative D genes whose sizes are compatible with the size of a surface antigen gene and which contain either pEDX or a sequence similar to it have been obtained by cloning. Are they the only members of the family? A Southern blot of total DNA digested with EcoRI has been hybridized with pEDX (Fig. 3). Only two bands (13 kb and 1.6 kb) are present, the intensity of the 1.6-kb band being greater than that of the 13-kb band. The 1.6-kb band corresponds to genes Dα and Dγ whereas Dβ gives a 13-kb band. Also, the same blot hybridized with the 7.8-kb EcoRI fragment ED7.8 of phage λD55 (Fig. 2) gives a group of bands 4.3-, 2.1- and 1.6-kb long that correspond to gene Dα, the 13-kb band to gene Dβ and the smeared bands that extend from 7.5 to 9 kb and around 12 kb to gene Dγ. Various probes used on Southern blots of total *Paramecium* DNA digested with different restriction enzymes always give a pattern of bands compatible with the restriction maps of the three genes described above (results not shown).

Expression of the D genes. To study the expression of these genes, we needed probes specific for each of the three putative mRNA. pEDX (and also S1, which is identical to pEDx) has been entirely sequenced. Partial DNA sequencing of the corresponding sequences (S2 and S3) of the two other genes was carried out and the sequences compared with that of pEDX (data not shown). The similarities between these sequences are extremely high (92% for S1 and S3 and 97% for S1and S2) in the part of the sequences corresponding to the 3′ end of the coding sequence, which should be present on a putative mRNA if it is expressed. However, three short regions of 23- to 25-bp

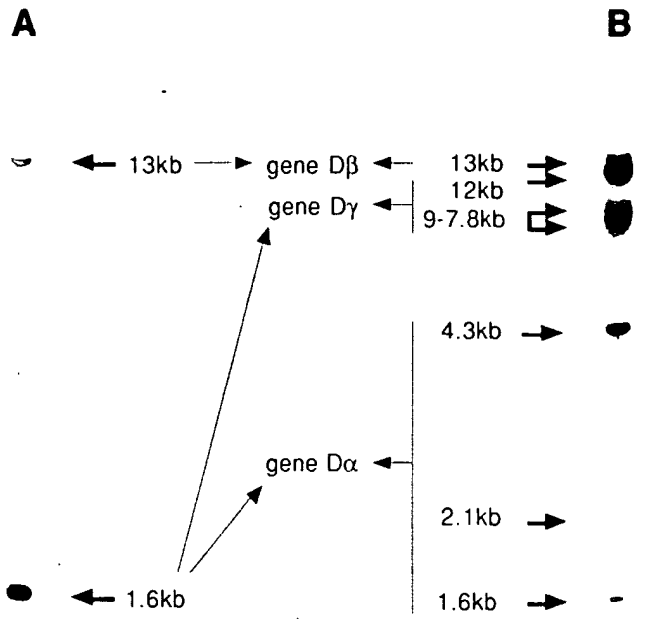

Fig. 3. Analysis of genomic DNA with D probes. Southern blots of EcoRI-cut *Paramecium primaurelia* strain 156 genomic DNA hybridized with probe pEDx (**A**) or probe ED7.8 (**B**). The various fragments indicated by thick arrows have sizes compatible with the three genes whose maps are shown in Fig. 2. The attribution of each fragment to the corresponding gene is shown by thin arrows.

have been chosen for their relatively high number of mismatches between one gene and the other two: Osr1 from S1 (gene Dα) is 23-bp long and has a two-mismatch difference with the other two sequences, S2 and S3 (Fig. 4), which are identical in this region. The same was obtained for S2 and S3 (genes Dβ and Dγ) with Osr2 and Osr3 (Fig. 4). The corresponding oligonucleotides were synthesized and used as probes on Northern blots of total RNA of D expressing cells. Care was taken to use stringency conditions suitable for exclusive hybridization of each oligonucleotide with the corresponding identical sequence. The results shown in Fig. 5 indicate without any ambiguity that only the mRNA of gene Dα is present in the D serotype.

Sequences of gene Dα and of the corresponding putative protein. As shown in previous articles, the coding sequence of these surface antigen genes can be determined accurately from the DNA sequence by the sudden drop in AT percentage when entering the coding sequence (and rise when leaving it) [7, 27]. It can also be determined by the bias in favor of A or T at the third position of each codon. This gives rise to a 3-bp periodicity of AT percentage in the coding sequences. Such a periodicity is completely absent from the flanking noncoding sequences. A 14-kb EcoRI fragment from phage λD2 covering entirely gene Dα has been entirely sequenced (see the thick hatched line on the map of phage λD2 in Fig. 2). Various regions along genes Dβ and Dγ have been sequenced too and compared with the sequence of gene Dα: the boundaries of the coding sequences (start codon ATG and stop codon TGA) have been determined without ambiguity using the tests mentioned above for all three genes. In between, the similarities of the coding sequences with Dα are extremely high (95% for gene Dβ and 91% for gene Dγ on the average) but outside these coding regions the percentage of similarity drops except for a few short stretches of sequences in the 5′ and 3′ noncoding regions (see below).

The open reading frame of gene Dα is 7632-bp long and does not contain introns. This size is compatible with the experi-

		0 10 20	number of mismatches
Osr1	S1	ATGGTTAATGTTATGACTATGAC	0pb
	S2	ATGGaTAATGTTATGAtTATGAC	2pb
	S3	ATGGaTAATGTTATGAtTATGAC	2pb
Osr2	S2	CAATCTTAATAATTCAAACGGAA	0pb
	S1	CAATCcaAATAATTCAAAtGGAA	3pb
	S3	CAATCcaAATAATTCAAAtGGAA	3pb
Osr3	S3	TTCAATCTAAAGCTGGACCATGTCT	0pb
	S1	TTaAATCaAAAaCTGGACCATGcCT	4pb
	S2	TTaAATCaAAAaCTGGACCATGcCT	4pb

Fig. 4. Oligonucleotides specific of each D gene. The sequences of oligonucleotides Osr1, Osr2 and Osr3 are shown and aligned with the corresponding sequences of the other two genes. Osr1 in region S1 is 23 bases long and the corresponding sequences in S2 and S3 present two mismatches (9%) noted in little bold prints. The oligonucleotides Osr2 (23 bases) and Osr3 (25 bases) in regions S2 and S3 show three base (13%) and four base (16%) differences, respectively, compared with the other two sequences. (Conditions for the specific hybridization of these oligonucleotides to the exact complementary sequence are given in the text.)

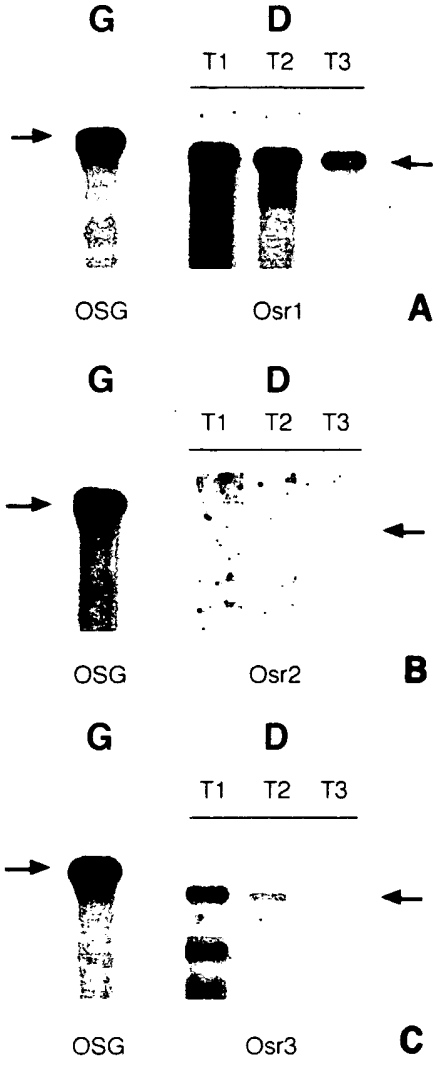

Fig. 5. Determination of the expressed D gene. Northern blots of total RNA of G- or D-expressing paramecia. The positions of the G or D mRNAs are indicated by arrows. On the left, G total RNA is hybridized with oligonucleotide OSG specific for G mRNA (OSG is a sequence at the 3' end of the coding sequence of the G gene: 5'-ACGATGACCCTTAATCGTAGACGATGTTTATTATTA-3') and used as a positive control for hybridization. On the right, the same experiment is done with D total RNA and the three oligonucleotides shown in Fig. 4: i.e. Osr1 in **A**, Osr2 in **B** and Osr3 in **C**. For each oligonucleotide, the membranes are washed at three temperatures (T1 = 45° C, T2 = 50° C and T3 = 55° C) and exposed for the same length of time (22 h).

mental size of the mRNA determined from the Northern blot of Fig. 1 (about 7500 bases). The coding sequence of the 156G gene of *P. primaurelia* is 8145-bp long. The 500-bp difference correlates nicely with the molecular weight difference of the mRNA.

From the nucleotide sequence of gene 156Dα, the amino acid sequence of the putative protein has been determined and is displayed in Fig. 6. The protein is 2543 amino acids long and has a calculated molecular mass of 267 kDa, which is close to the experimental value, suggesting that the extent of the maturation process of this protein is small. The amino acid composition was deduced from the amino acid sequence using the special deviated genetic code of *Paramecium* [7].This protein is rich in cysteine (10.5%), threonine (13.5%) and serine (9.1%), a property shared with all other sequenced surface antigens [25–27, 36]. Because the amino acid composition of the 156Dα protein has not yet been determined experimentally, we compared the deduced amino acid composition with the experimental amino acid compositions of two alleles of the 156Dα gene: 90D and 178D [16]. This is possible because the amino acid sequences of two alleles of a surface antigen gene are extremely similar (for instance the G genes from strains 156 and 168 have a 93% similarity, see [26]) and because the amino acid compositions of all sequenced surface antigens are extremely homogeneous [27]. A χ^2 test was performed to compare the amino acid composition of the putative 156Dα protein and an average of the two experimental amino acid compositions of 90D and 178D. The contribution to the χ^2 value of each amino acid is low (between 0.2 and 1.9) except for serine and glycine. The same discrepancy for serine and glycine was previously observed in the 156G protein [27].

The amino acid sequence of the 156Dα protein displays the usual pseudoperiodicity typical of *Paramecium* surface antigens (156G: [27] and 168G: [26] of *P. primaurelia*; 51A and 51C: [25], 51B: [36], and 51Dα: see accompanying paper, of *P. tetraurelia*). The position of the cysteine residues are well conserved and are aligned vertically in Fig. 6 to enhance this effect. However, in 156Dα, the pseudoperiodicity appears less regular than in 156G, 168G and 51A and the vertical alignment in Fig. 6 is less clear [25–27]; also, the distance between two cysteine residues is sometimes extremely large (line 8) and some half periods contain three, five or seven cysteine residues (lines 8, 27, 28 and 29) instead of four in the three above mentioned proteins. As in 51C protein [25], a remarkable feature is the

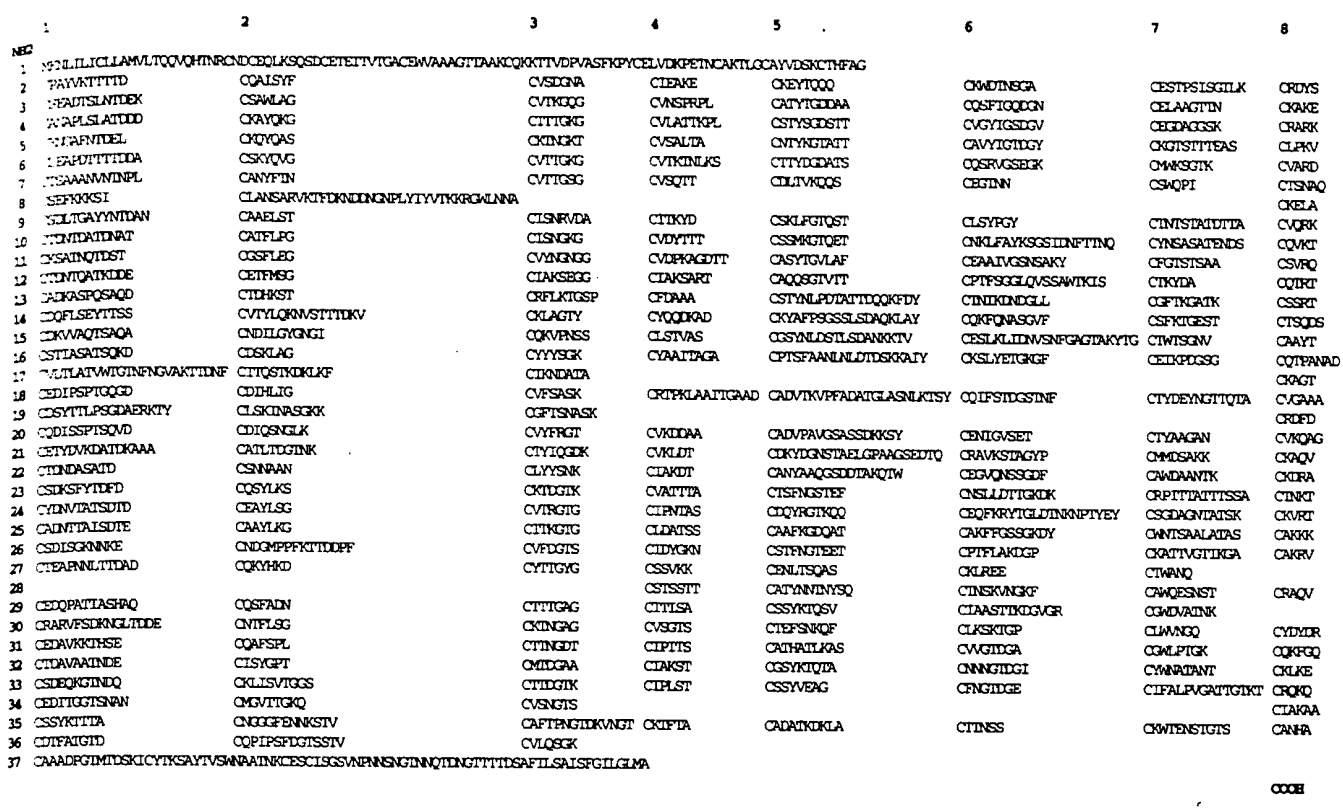

Fig. 6. Pseudoperiodic structure of the 156Dα protein. The complete amino acid sequence is represented in such a way that each line corresponds to one period, as in Prat et al. [27]. Each period contains eight cysteine residues, except for the two ends and five incomplete periods (lines 8, 17, 19, 28 and 34). The database accession number of the 156Dα gene is a562F34s (X96616).

absence of central repeats, which were observed in 156G, 168G, 51A and 51B proteins [25–27, 36]. This again strengthens the point that apparently two kinds of surface antigen structures exist: with or without central repeats. In Fig. 7, matrices of identity between the 156D protein and various other surface antigen sequences are displayed: in a, the comparison between the 156Dα and itself shows the absence of central repeats but the presence of a repeated motif at both ends of the protein. The latter appears to be a distinctive feature of the 156Dα protein; in b, two D surface protein sequences, 156Dα of *P. primaurelia* and 51Dα of *P. tetraurelia* (see accompanying paper) are compared: the continuity of the diagonal illustrates the similarity (82%) of the two sequences. In C and D, the comparison with two surface antigen sequences with (156G: [27]) or without (51C: [25]) central repeats shows that in both cases, the sequences are similar except for a central part representing roughly one third of the sequence. This study and the previously published ones show that all these surface protein sequences are most similar at both ends and can display large sequence variations in the central part [25, 26, 36].

Comparison of the 5′ and 3′ noncoding sequences. In Fig. 8, 3′ and 5′ noncoding sequences of various surface antigen genes upstream of the initiator ATG and downstream of the TGA stop codon are aligned. As mentioned by previous authors three consensus sequences, at −10 and −60 upstream of the ATG codon, and one 30-bp downstream of the TGA stop codon, present in all surface antigen genes, are also present in the three D genes studied here [32]. No motif common to a subset of these sequences has been found. Also, these motifs are absent in other *Paramecium* genes already sequenced [10, 18]. This suggests that these common motifs could be binding sites for transcription factor(s) necessary for general (and not specific) surface antigen gene expression.

A region of variable DNA rearrangement is present at the 5′ end of gene Dγ. As previously mentioned (Fig. 2), a region of variable size is present at the 5′ end of gene Dγ, which is most likely to arise from variable DNA rearrangements that occur during macronuclear development. Six versions are represented by the six recombinant phages λD15, λD81, λD19, λD57, λD22 and λD55 and the variable region is contained in a Hc*–EcoRI fragment (Fig. 2). To determine the nature of these rearrangements, this Hc*–EcoRI fragment has been sequenced for each of the six recombinant phages. In Fig. 9, the sequences have been aligned vertically when they are rigorously identical. At first glance all these sequences appear to share large regions of identity, but some subsequences present in recombinants are absent in others as if they had been eliminated. The micronuclear DNA corresponding to these macronuclear regions has not yet been cloned. Therefore, the only possible investigation we have done is to order the subsequences of these six macronuclear versions as they are in the macronucleus. For instance, the segment SG1 (Fig. 9) of 200 bp, which is present only in λD57, could be located at various positions in macronuclear DNA and not necessarily at the position shown in Fig. 9. However, this position appears to be correct since PCR products amplified from genomic DNA using O1 and O2 as primers (Fig. 9) do not hybridize with O4, which is contained in SG1, suggesting that O4 is effectively at the right of O1 and O2. Moreover, PCR amplification of macronuclear DNA with oligos O3 and O6 (Fig. 9 for their location) reveals the presence of smeared bands that

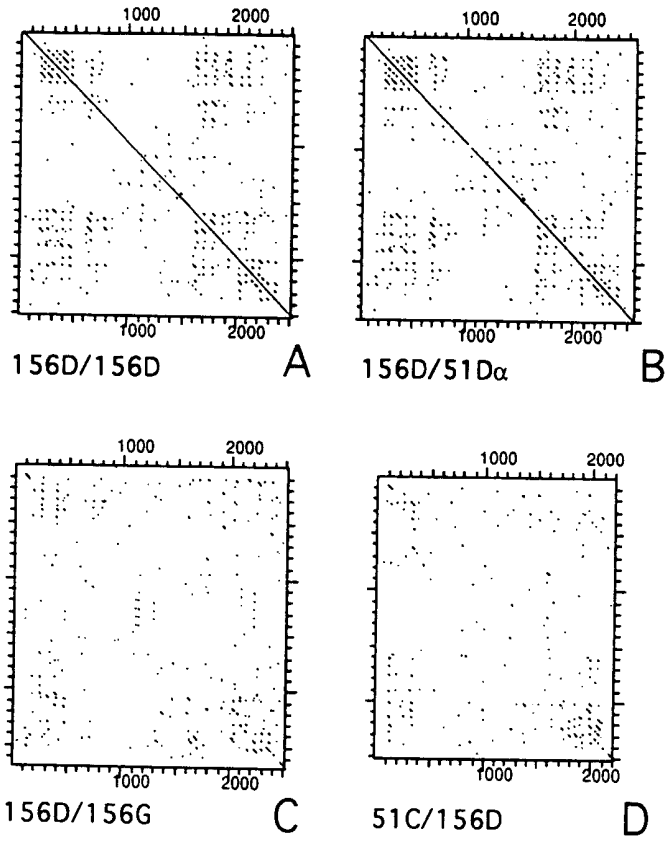

Fig. 7. Dot matrix comparison of the 156Dα deduced amino-acid sequence with other surface antigen sequences. **A.** 156Dα versus 156Dα. **B.** 156Dα versus 51Dα. **C.** 156Dα versus 156 G. **D.** 156Dα versus 51C. DNA and protein sequences were analyzed using the program "DNA Strider" [19]. The accession number of 156Dα gene is a562F34s (X96616).

A

```
     -200                                                        -151
156dα  ...CAACATC CTTGTTCTTT GCATTAGTAT AATGAACATA ATATAATATT
156dβ  ......... ......... ......... ......... .........
156dγ  ......... .....CCTT GTTTTAGATT GATCTAGAAA AACAAAAGTT
51dα   ...TACCTTA TCAAACCGAC TTTTTAGTGC CTTCGTAAAT TGAAATATAA

       -150                                                        -101
156dα  TATAAATATT TATGATGTGG ATTAATTATT CAAATTAAAT TCTTGTCAAA
156dβ  ......... ......... ......... ......... .........
156dγ  AAAAATATTA ATTATGTCAA TTAAtTTTCA ATTTTAATTC TTAATCAACT
51dα   AGCAAAAATT ATATTAATTA ATTATAAGAA TTAAAATGAA ATCTAAATTA

       -100                                                        -51
156dα  ACATTAAAAA AGATTATTAT GTAATTAATT AAATAAAAAA AATTAATTTT
156dβ  ......... ......... ......... ......... .........
156dγ  TTTTTTTAAG ATTTAGTATG CAATTATCTA AGCAAAAAAA ATTAATTTTT
51dα   TGAATAATTA ATTTGGATAA AAATATTATC ACATTAATAT CTTAAAAAAA

       -50                                                         -1
156dα  TTGAGATTAT TTAGCTTCCC TAACTTTAAT ATATTCAATT AATACTTTTA
156dβ  ......... ......... ......... AATTTCAATT AATACTTTTT
156dγ  TGAAATTATT TAACCTCCCT TTATCTAATA TTTTTTATTT AATACTTTAA
51dα   ATTAAAATTT CCCTAAATAA AACCTTTTTC CTTATAAATG TAATTTTAAT
```

B

```
       1                                                           50
156dα  TGATTTTGAT TTTCATATTA TATAGGGAAA ATTTTAATAT TACAAATTAT
156dβ  TGATATTGAT TTTCATATTA TATAGGGAAA ATATTAATAA TACAAATTAT
156dγ  TGATTTTGAT TTTCATTTTA TAGGGAAAAT ATTAATATTA CAAGTTATTT
51dα   TGATTTTGAT TTTCATATTA TATATAGGGA AAAATTTATT AATGAAAAAA

       51                                                          100
156dα  TTTTATCTCT CAATAGAGGG TTTTTTAATA AATATAAAAC TCTTAAAATT
156dβ  TTATCTCTTA ATAAGGGTGC TTTAGTAATA TTAAA..... .........
156dγ  TATCTCTTAA TAGAGGTGTT TTAGTAGATG TTAAATTCGT AAAATTAATA
51dα   ATTAATTTAA TCTTTAAATA AGATTGTTTC AAAAATAAAA TCTAAATCAT
```

Fig. 8. Comparison of 5′ and 3′ noncoding sequences. Noncoding sequences in the 5′ (**A**) and 3′ (**B**) part of 156Dα, Dβ, Dγ and 51Dα genes were aligned using the GCG program. **A.** Position −1 is immediately before the translation initiation codon. **B.** Position +1 is the first stop codon. The conserved sequences are underlined. The accession numbers are: a562F34s (X96616)—156Dα upstream and downstream noncoding sequences and coding sequence; R436K87I (X96626) and o456b63N (X96627)—156Dβ 5′ and 3′ noncoding sequences, respectively; u518a81y (X96629) and F478U10W (X96628)—156Dγ 5′ and 3′ noncoding sequences, respectively.

hybridize with SG1. This shows that SG1 is between O3 and O6 but also that in the macronucleus the rearrangements are much more complex than what is represented by those six phages. This high degree of complexity is also revealed at various other positions in this region; for instance PCR amplification using O3 and O4 as primers also gives multiple discrete bands. However, no PCR product was obtained with O4 and O5, showing that SG2 is at the right of SG1(Fig. 9). Indeed, PCR amplification with O5 and O6 gives a unique band of 340 bp as in phage λD22.

The fact that the sequences in this region, when common to multiple phages, are identical, strongly suggests that they arise from the same micronuclear sequence by alternative elimination of internal sequences called IES (for a review, see [33]). Also, in favor of a variable elimination of different IES is the presence of a 5′-TA-3′ dinucleotide at each interruption shown by a dotted line in Fig. 9 (results not shown). The sequences from each recombinant phage bordering the SG3 segment of 130 bp present in λD15, λD19 and λD55 but absent in λD81 are shown in Fig. 10A. The comparison of λD81, where the 130-bp sequence is absent, and of the other three (λD15, λD19 and λD55) where it is present, shows the presence of a 5′-TA-3′ repeat at both ends of the eliminated sequences. The fact that 5′-TA-3′ is repeated twice in λD55 and only once in λD15 and λD19 can be explained by the model shown in Fig. 10B where two left IES boundaries are alternatively used. A complete description of the DNA rearrangement will have to await the cloning of the corresponding micronuclear copy.

Caryonidal variation of the variable DNA rearrangement. In many cases, variable DNA rearrangements in *Paramecium* display a caryonidal variation [6, 17]. This is also true for this region: macronuclear DNA from different caryonidal clones cut with EcoRI are hybridized on Southern blots (Fig. 11) with the BHc probe shown in Fig. 2 and covering the 5′ end of gene Dγ at the border of the variable DNA rearrangement region. A 12-kb band represented by λD15 and a smeared 9- to 7.5-kb band represented by the other five (λD81, λD19, λD57, λD22 and λD55) are present in all caryonides, but with different intensities. More precisely, a low intensity of the smeared band correlates with a higher intensity of the 12-kb band. This again supports the idea that these macronuclear versions arise from the same micronuclear locus and that the macronuclear versions giving rise to the 12-kb EcoRI band could be versions where the right

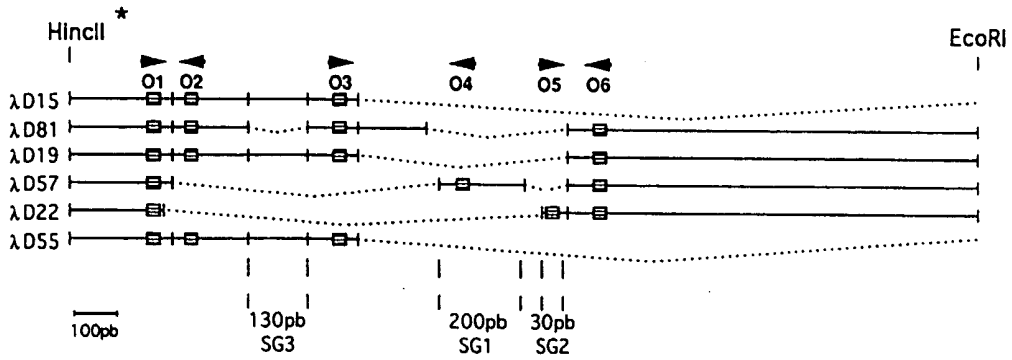

Fig. 9. Map of the variable region. Alignment of the sequences to the right of the Hc* site (Fig. 2) of the six representative phages λD15—accession number B545F67o (X96630); λD81—accession number p590L20n (X96635); λD19—accession number Q566R53h (X96631); λD57—accession number E587q13S (X96634); λD22—accession number F572A59u (X96632); λD55—accession number G579V15t (X96633). Vertical bars indicate positions where sequence identity is interrupted. Unique sequences SG1 and SG2 were ordered by PCR using oligonucleotides O1, O2, O3, O4, O5 and O6 (boxed areas).The 5′–3′ orientation of the primers is indicated by an arrow. The sequences of the EcoRI extremity of λD81, λD19, λD57, λD22 are identical, but different from the λD15 and λD55 sequences (not shown), which also differ from each other.

EcoRI site present in the others is absent due to sequence elimination. No attempt has been made to determine the overall extension of this region of alternative DNA rearrangements.

DISCUSSION

A subfamily of three surface antigen genes of the D type, Dα, Dδ and Dγ, has been found in strain 156 of *Paramecium primaurelia*. Only one (Dα gene) is expressed in the D serotype and it corresponds to the major high molecular weight mRNA species. A minor species in the same molecular weight range, but with a slightly lower electrophoretic mobility, has been detected. It could be another surface antigen mRNA coexpressed with the D surface antigen, but in this case the failure to cross-hybridize with the 3′ part of its sequence, even at low stringency, with G or D surface antigens probes suggests that it probably belongs to another family. However, if this minor species was the mRNA of a surface antigen, this would not be surprising. Indeed, similar cases of co-transcription of surface protein RNA have been described in the literature; for instance, in *P. tetraurelia*, the mRNA of the C surface protein is present in the cytoplasm of cells expressing the H serotype, but the C protein is not detected at the cell surface [14]. Also, P. Margolin [20] has shown that in the strain 172 of *P. tetraurelia*, surface antigen M is often weakly coexpressed with D. That the minor species could be the mRNA of a surface antigen is reinforced by protein analysis [4] of D expressing cells. In the high molecular weight range, a major band corresponding to the major mRNA can be seen along with a minor band migrating slightly more slowly that could correspond, based on its relative intensity, to the minor mRNA species. Both are membrane proteins since they immunologically react with an anti-CRD monoclonal antibody.

The sequence of the 156Dα protein is given in Fig. 6. It shows the same structural features as those of previously published surface antigen sequences, namely a highly periodic structure revealed by the regular position of eight cysteine residues per period [27]. Two types of surface antigen gene structures have been described in the literature: one that contains central repeats like 156G and 168G in *P. primaurelia* [26, 27] or 51A [25] and 51B [36] in *P. tetraurelia*, and another that does not contain these repeats, making the amino acid sequence slightly shorter,

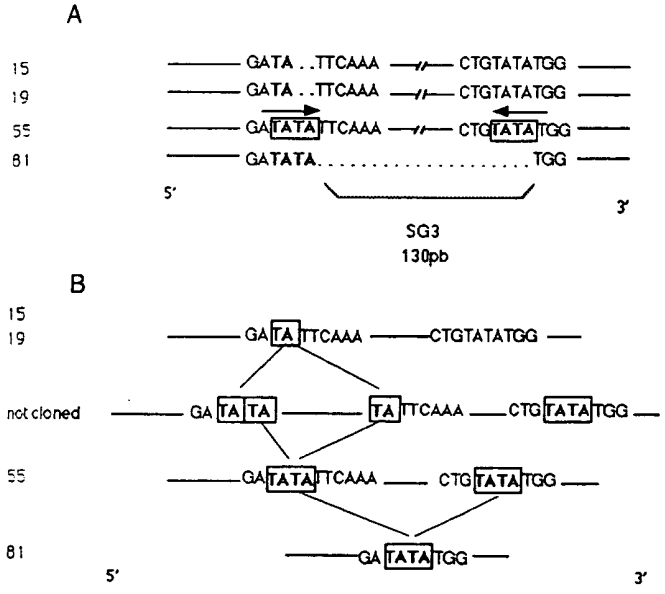

Fig. 10. Sequences of λD15, λD19, λD55 and λD81 in the variable region. The sequence of SG3 junctions is shown in **A**. The direct repeats 5′-TATA-3′ on each side of SG3 in λD55 are framed and a 5-bp palindromic sequence is indicated by the two arrows. The dinucleotide 5′-TA-3′ is missing in λD15 and λD19 so that the direct repeat is only 5′-TA-3′. In λD81 the SG3 130-bp segment is absent and one of the direct repeats 5′-TATA-3′ is retained. Continuous lines symbolize identical sequences. A possible interpretation for the observed polymorphism at the left SG3 junction is shown in **B**. Two elimination processes from a noncloned version of this region can explain the presence either of a 5′-TA-3′ in λD15, λD19 or of 5′-TATA-3′ in λD55.

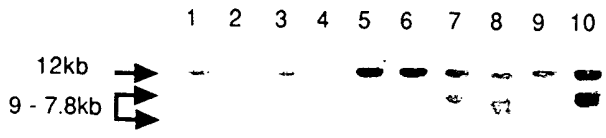

Fig. 11. Caryonidal analysis of the variable region. A Southern blot of EcoRI-restricted genomic DNA from different caryonidal cultures (1–10) of *P. primaurelia* strain 156, was hybridized with probe BHc, which is common to all six cloned versions in the 5′ region of the Dγ gene (Fig. 2). The arrows indicate the position of the 12-kb band and the 7.5- to 9-kb smear. Samples 1 to 10 are DNA.

such as 51C [25]. In this respect, the 156Dα protein belongs to the second category, but the resemblance stops there since the matrix of identity between 51C and 156D displayed in Fig. 7D shows only a few randomly scattered stretches of identity. The sequences of the D surface proteins of *P. primaurelia* and *P. tetraurelia* show remarkable similarity (Fig. 7B), a feature that has been already noticed between surface antigens of the two varieties: for instance, A from *P. tetraurelia* and G from *P. primaurelia*. Moreover, two regions at the NH_2 and at the COOH termini occupying a rather symetrical position in the two D sequences clearly appear as being made of a repeated motif as shown by the line segments equidistant and parallel to the first diagonal (see the identity matrices of Fig. 7A, B). Besides, these two motifs appear to be similar as shown by the enhancement of similarity segment numbers in the upper right corner of the same figure.

We have shown that only three genes belong to the D subfamily (Fig. 2, 3). The other two nonexpressed D genes, Dβ and Dγ, are extremely similar to the expressed Dα gene at least along the parts that have been sequenced. Also along those same sequences, their open reading frames are not interrupted by STOP codons, suggesting that they are not pseudogenes. These three macronuclear genes must arise from three different micronuclear genes and not from an alternative processing of a unique micronuclear gene since they differ by numerous point mutations all along their sequences. Two interpretations of the existence of these three very similar surface antigen genes can be mentioned at this stage. Firstly, these three genes could have been obtained by recent duplications processes and indeed two of them, Dβ and Dγ, are physically closed to each other. In this respect, the three genes would represent isoforms of the D subfamily. This would be a new and remarkable feature of the surface antigen gene properties since for each of the others characterized serotypes, a unique gene have been cloned, in perfect agreement with genetic data showing these genes to be in unique copy in the genome (for a review, see [29]). Secondly, Dα would be the unique gene of the D serotype and the two Dβ and Dγ genes could be assigned to two others serotypes closely related to D. They could be the analogs in *P. primaurelia* of surface antigens J and M, which have been characterized in *P. tetraurelia* as immunologically related to D [28]. It is worth mentioning here that two surface antigens can have very homologous sequences without belonging to the same serotype: a good example is given by the two alleles 156 and 168 of the G surface antigen; the determinants of the serotypes have been shown to lie within the central repeats, which are the only part of the two sequences that differs significantly [8]. These central repeats represent a short amino acids motif (74 amino acids for 156 and 73 for 168) compared to the length of the whole sequence (2715 amino acids for 156 and 2704 for 168). Apart from this example, we have a complete ignorance of what in a surface antigen amino acid sequence determines the serotype. Therefore the genetic data mentionned above simply shows that the D serotype corresponds to the expression of the Dα gene alone.

The noncoding sequences of genes Dα, Dβ and Dγ, either at the 5′ or at the 3′ end, are very similar (Fig. 8). In cases where the expression of surface antigens has been tested, primarily by microinjection of a plasmid containing a defined surface antigen gene, not only the expression of this surface antigen but also the regulation of its expression as a function of temperature has been found to be controlled by sequences within the gene itself or within a few hundred base pairs at both 5′ and 3′ ends [15, 21, 22]. The fact that the coding sequences of genes Dα, Dβ and Dγ and the 5′ and 3′ noncoding sequences are very similar, but that only one of these three genes, is expressed indicates that some subtle changes in their sequences may control expression. Microinjection of the two nonexpressed genes, Dβ and Dγ into the macronucleus will be of great interest.

A region of variable sequence has been identified upstream of the Dγ gene and the sequences of six versions of this region have been partially determined. Their structures strongly suggest that they are derived from the same micronuclear copy by alternative DNA sequence elimination. Indeed, the comparison of the sequences shows the systematic presence of a 5′-TA-3′ at each point where different versions start diverging in their sequence. Elimination of IES has already been reported by different authors in both *P. primaurelia* and *P. tetraurelia*, and of the two 5′-TA-3′ dinucleotides that border the IES only one is conserved after elimination [1, 30, 37; 39]. Therefore, alternative IES elimination could account for the results obtained. However, we have cloned only six versions that have been partially sequenced and Southern blots of this region show a large heterogeneity of variable versions; this suggests a much larger number of possible versions. Also, the extent of this variable macronuclear region has not yet been determined.

In the absence of the cloned micronuclear copy it appears difficult to build up a model of DNA processing that would justify the maintenance of one or two 5′-TA-3′ in some macronuclear versions as shown in Fig. 10A; but at least three possible explanations can be considered: an alternative elimination of a family of IES either consecutive in the micronuclear genome or imbricated as in Fig. 10B, a microheterogeneity in the choice of boundaries, and finally, a microheterogeneity in the junction created by IES excision that would originate from the mechanism of excision itself (a variable DNA repair before religation, for instance) and would be reminiscent of some transposon excision [40]. The functional role of this region (if any) has not yet been determined, but it is tempting to consider that it might be involved in the regulation of Dγ expression or in the yet unknown mechanism generating variability of these surface antigens.

ACKNOWLEDGMENTS

We wish to thank Laurence Amar, Anne-Marie Keller, Anne Le Mouël and Helmut Schmidt for stimulating discussions, and Linda Sperling and Eric Meyer for critically reading the manuscript. This work was supported by grant no. 90261 from the Ministère de l'Enseignement Supérieur et de la Recherche, grant from Association pour la Recherche sur le Cancer and grant no. 30 from the Groupement de Recherches et d'Etudes sur les Génomes.

LITERATURE CITED

1. Amar, L. 1994. Chromosome end formation and internal sequence elimination as alternative genomic rearrangements in the ciliate. *J. Mol. Biol.*, **236**:421–426.

2. Capdeville, Y. 1971. Allelic modulation in *Paramecium aurelia* heterozygotes. Study on G serotype syngen 1. *Mol. Gen. Genet.*, **112**: 306–316.

3. Capdeville, Y. 1979. Intergenic and interallelic exclusion in *Paramecium primaurelia*: immunological comparisons between allelic and nonallelic surface antigens. *Immunogenetics*, **9**:77–95.

4. Capdeville Y., Caron, F., Antony, C., Deregnaucourt, C. & Keller, A. M. 1987. Allelic antigen and membrane-anchor epitopes of *Paramecium primaurelia* surface antigens. *J. Cell Sci.*, **88**:553–562.

5. Capdeville, Y., Deregnaucourt, C. & Keller, A.M. 1986. Immunological evidence of a common structure between *Paramecium* surface antigens and *Trypanosoma* variant surface glycoproteins. *Exp. Cell Res.*, **161**:495–508.

6. Caron, F. 1992. A high degree of macronuclear chromosome polymorphism is generated by variable DNA rearrangements in *Paramecium primaurelia* during macronuclear differenciation. *J. Mol Biol.*, **225**:661–678.

7. Caron, F. & Meyer, E. 1985. Does *Paramecium primaurelia* use a different genetic code in its macronucleus? *Nature*, **314**:185–188.

8. Caron, F. & Ruiz, F. 1992. A method for the amplification of *Paramecium* micronuclear DNA by polymerase chain reaction and its application to the central repeats of *Paramecium primaurelia* G surface protein. *J. Protozool.*, **39**:312–318.

9. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry*, **18**:5294–5299.

10. Cohen, J., Dupuis, P. & Viguès, B. 1990. Expression of ciliate gene in *E. coli* using a suppressor tRNA to read the UAA and UAG glutamine codons. *J. Mol. Biol.*, **216**:189–194

11. Devereux, J., Haerberli, P. & Smithies, O. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucl. Acids Res.*, **12**:387–395.

12. Dryl, S. 1959. Antigenic transformation in *Paramecium aurelia* after homologous antiserum treatment during autogamy and conjugation. *J. Protozool.*, **6**(suppl.):25.

13. Emilsson, V. & Kurland, C. G. 1990. Growth rate dependence of transfer RNA abundance in *Escherichia coli*.. *EMBO J.*, **9**:4359–4366.

14. Gilley, D., Rudman, B., Preer, J. R., Jr. & Polisky, B. 1990. Multilevel regulation of surface antigen gene expression in *Paramecium tetraurelia*. *Mol. Cell. Biol.*, **10**:1538–1544.

15. Godiska, R., Aufderheide, K. J., Gilley, D., Hendrie, P., Fitzwater, T. L., Preer, L. B., Polisky, B. & Preer, J. R., Jr. 1987. Transformation of *Paramecium* by microinjecting a cloned serotype gene. *Proc. Natl. Acad. Sci.* USA, **84**:7590–7594.

16. Jones, I. G. 1965. Studies on the characterization and structure of the immobilization antigens of *Paramecium aurelia*. *Biochem. J.*, **96**: 17–23.

17. Keller, A. M., Le Mouël, A., Caron, F., Katinka, M. & Meyer, E. 1992. The differential expression of the G surface antigene alleles in *Paramecium primaurelia* heterozygous cells correlates to macronuclear rearrangement. *Dev. Genet.*, 13:306–317.

18. Kink, J. A., Maley, M. E., Preston, R. R., Ling, K.-Y., Wallen-Frienman, M. A., Saim, Y. & Kung, C. 1990. Mutation in *Paramecium* calmodulin indicate functional differences between the C-terminal and the N-terminal lobes in vivo. *Cell*, **62**:165–174.

19. Marck, C. 1988. "DNA Strider": a "C" program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. *Nucl. Acids Res.*, **16**:1829–1856.

20. Margolin, P. 1956. An exception to mutual exclusion of the ciliary antigens in *Paramecium aurelia*. *Genetics*, **41**:685–699.

21. Martin, L. D., Pollack, S., Preer, J. R., Jr. & Polisky, B. 1994. DNA sequence requirements for the regulation of immobilization antigen A expression in *Paramecium aurelia*. *Dev. Genet.*, **15**:443–451.

22. Meyer, E. 1992. Induction of specific macronuclear developmental mutations by microinjection of a cloned telomeric gene in *Paramecium primaurelia*. *Genes & Dev.*, **6**:211–222.

23. Meyer, E., Caron, F. & Baroin, A. 1985. Macronuclear structure of the G surface antigen gene of *Paramecium primaurelia* and direct expression of its repeated epitopes in *E. coli*. *Mol. Cell Biol.*, **5**:2414–2422.

24. Meyer, E., Caron, F. & Guiard, B. 1984. Blocking of in vitro translation of *Paramecium* messenger RNAs is due to messenger RNA primary structure. *Biochimie*, **66**:403–412.

25. Nielsen, E., You, Y. & Forney, J. 1991. Cysteine residue periodicity is a conserved structural feature of variable surface proteins from *Paramecium tetraurelia*. *J. Mol. Biol.*, **222**:835–841.

26. Prat, A. 1990. Conserved sequences flank tandem repeats in two alleles of the G surface protein of *Paramecium primaurelia*. *J. Mol. Biol.*, **211**:521–535.

27. Prat, A., Katinka, M., Caron, F., & Meyer, E. 1986. Nucleotide sequence of the *Paramecium primaurelia* G surface protein. A huge protein with a highly periodic structure. *J. Mol. Biol.*, **189**:47–60.

28. Preer, J. R., Jr. 1959. Studies on the immobilization antigens of *Paramecium*. IV. Properties of the different antigens. *Genetics*, **44**: 803–819.

29. Preer, J. R., Jr. 1986. Surface antigens of *Paramecium*. *In*: Gall, J. G. (ed.), The Molecular Biology of Ciliated Protozoa. Academic Press, New York. Pp. 301–339.

30. Preer, L. B., Hamilton, G. & Preer, J. R., Jr. 1992. The isolation of micronuclei from *Paramecium tetraurelia*; serotype 51A gene has internally eliminated sequences. *J. Protozool.*, **39**:678–682.

31. Preer, J. R., Jr., Preer, L. B. & Rudman, B. M. 1981. mRNAs for the immobilization antigens of *Paramecium*. *Proc. Natl. Acad. Sci.* USA, **78**:6776–6778.

32. Preer, J. R., Jr., Preer, L. B., Rudman, B. M. & Barnett, A. J. 1987. Molecular biology of the genes for immobilization antigens in *Paramecium*. *J. Protozool.*, **34**:418–423.

33. Prescott, D. 1994. The DNA of ciliated protozoan. *Microbiol. Rev.*, **58**:233–267.

34. Rackwitz, H. R., Zehetner, G., Frishauf, A. M. & Lehrach, H. 1984. Rapid restriction mapping of DNA cloned in lambda phage vectors. *Gene*, **30**:195–200.

35. Sambrook, J., Fritsch, E. F. & Maniatis, T. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor University Press, Cold Spring Harbor.

36. Scott, J., Leeck, C. & Forney, J. 1993. Analysis of the micronuclear B type surface protein gene in *Paramecium tetraurelia*. *Genetics*, **133**:189–198.

37. Scott, J., Leeck, C. & Forney, J. 1994. Analysis of the micronuclear B type surface protein gene in *Paramecium tetraurelia*. *Nucl. Acids Res.*, **22**:5079–5084.

38. Sonneborn, T. M. 1974. *Paramecium aurelia*. *In:* King. R. C. (ed.), Handbook of Genetics. Plants, Plant Viruses and Protists. Plenum Press, New York. **2**:469–594.

39. Steele, C. J., Barkocy-Gallagher, G. A., Preer, L. B. & Preer, J. R., Jr. 1994. Developmentally excised sequences in micronuclear DNA in *Paramecium*. *Proc. Natl. Acad. Sci.* USA, **91**:2255–2259

40. Van Luenen, H. G., Colloms, S. D. & Plasterk, R. H. 1994. The mechanism of transposition of Tc3 in *C. elegans*. *Cell*, **79**:293–301.

Received 10-9-95, 3-19-96; accepted 3-20-96